FY 2002 ANNUAL REPORT

SPECIALIZED INFORMATION SERVICES

FISCAL YEAR 2002 PROGRAMS AND SERVICES

Jack W. Snyder, M.D., Ph.D.
Associate Director

The Toxicology and Environmental Health Information Program (TEHIP), known originally as the Toxicology Information Program, was established 35 years ago within the National Library of Medicine in the Division of Specialized Information Services (SIS). Over the years TEHIP has provided for the increasing need for toxicological and environmental health information by taking advantage of new computer and communication technologies to provide more rapid and effective access to a wider audience. We have moved beyond the bounds of the physical NLM, exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever these sources may reside. Resources include chemical and environmental health databases and Web-based information resource collections. Development of HIV/AIDS information resources became a focus of the Division several years ago, and now includes several collaborative efforts in information resource development and deployment, including a focus on the information needs of other special populations.

The SIS Web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server (http://sis.nlm.nih.gov) users can access interactive retrieval services in toxicology and environmental health, HIV/AIDS information, or special population health information; find program descriptions and documentation; or be connected to outside related resources. Continuous refinements and additions to our Web-based systems are made to allow easy access to the wide range of information collected by this Division. Our usage has continued to increase over the past year with access to all toxicology and HIV/AIDS data free over the Internet.

In FY2002 SIS focused on several projects for significant re-engineering and initiated several new opportunities to enhance SIS information resources and provide new services in emerging areas. Prototypes have been developed which utilize graphical display of data from our information resources, innovative access and interfaces for consumers, and geographical information systems. Highlights for 2002 include a new user interface and multi-database search capability for TOXNET, our premier collection of databases on toxicology, hazardous chemicals, and related areas; Haz-Map, an occupational toxicology database designed to link jobs to hazardous job tasks that may be associated with occupational diseases and their symptoms; ALTBIB, an improved search of the multiple bibliographies of Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing; several new Toxicology & Environmental Health Special Topic Web resource pages, including Chemical Warfare Agents and Arsenic; TOXMAP, a prototype system that uses maps of the U.S. to help users visually view data about chemicals released into the environment and easily connect to related environmental health information; ToxTown, a graphical portal to chemicals you might encounter in everyday life, in everyday places; successful installation of four PAHO/NLM Disaster Preparedness Information Centers in Honduras and Nicaragua, with the addition of El Salvador this year; expanded Native American outreach initiatives; and new minority outreach activities with the Historic Black Colleges and Universities, United Negro College Fund Special Projects, and the National Medical Association.

Resource Building

The wide range of resources related to toxicology and environmental health information, HIV/AIDS information, and special populations information include many databases that are created or acquired as well as other services and projects.

Haz-Map database was released in 2002 on the Internet (http://hazmap.nlm.nih.gov). It is an occupational toxicology database designed to link jobs to hazardous job tasks that are linked to occupational diseases and their symptoms. It is a relational database of chemicals, jobs, and diseases. The Haz-Map jobs table is based on the 1997 Standard Occupational Classification system. The

industries table is based on the Standard Industrial Classification system. The diseases table is based on the International Classification of Diseases (ICD-9). Information from textbooks, journal articles, and electronic databases (HSDB, ACGIH Documentation of TLVs, ATSDR Toxicological Profiles, NIOSHTIC, and others) was classified and summarized to create the database. A user may search this occupational database by chemical agent, occupational disease and by job type.

ChemIDplus (Chemical Identification File) is an NLM online chemical dictionary, which contains over 360,000 records, primarily describing chemicals of biomedical and regulatory importance, and available on the Internet (http://chem.sis.nlm.nih.gov/chemidplus). ChemIDplus features include chemical structure search and display for over 143,000 chemicals, and hyper linked locator fields that retrieve data for a given chemical from other resources such as TOXLINE, MEDLINE or HSDB as well as EPA and ATSDR. Over 15,000 records of regulatory interest collectively known as SUPERLIST are also available and hyper linked in ChemIDplus. During FY2002 over 54,000 queries per month were made of this database. The database was enhanced by the addition of a variety of new locators pointing to international resources, including coverage of agents found in ClinicalTrials.gov. All drug data from the publication USPDDN from the United States Pharmacopeial Convention, Inc. was also reloaded. A new Web-based maintenance system named DBMaint2 was developed and tested, and will be available in 2002. It integrates text and structure input and modification, which will increase efficiency by staff and contractors. In addition, a new chemical spell checker was developed and tuned for ChemIDplus data. This will be released in 2002, and will help users to retrieve substances by chemical name, a type of data that is highly susceptible to spelling errors by users. In FY2002, a new test version of ChemIDplus was developed for release in 2003.

The **Hazardous Substances Data Bank (HSDB)** continues to be a highly used resource, averaging over 48,000 searches each month (a 20% increase over FY2001). Increased emphasis continues to be placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with past recommendations of the Board of Regents' Subcommittee on TEHIP. In 2002, there has also been an increased emphasis on adding to HSDB new chemicals with the potential for high

toxicity and high human exposure. Over 125 new chemicals were added in 2002, including new pesticides and environmental pollutants. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,671 HSDB records. Because of increased staff efforts, more records are being processed through special enhancements, including source updates from various peer-reviewed files. Special summary information is being prepared to allow easier presentation of information at a health consumer level. The process of developing a new Web-based system for HSDB creation, review, and maintenance is continuing. This will replace the current Remote Data Entry System (RDES) next year. As part of this effort, a test version of a relational HSDB database using the MySQL database application was created, and a new client-server interface was programmed to allow easier updates.

The Toxicology Data Network (TOXNET), NLM's information system providing database management for many of its toxicology files, has moved from a networked microprocessor environment to a UNIX-based platform (Solaris Version 2.6) on a SUN Enterprise 3000 computer. Integration of this configuration with other SIS database creation systems and the Web access to them is currently underway. In FY2002, SIS introduced a new search interface to allow integrated access to the SIS toxicology and environmental health databases. This new search interface (http://toxnet.nlm.nih.gov) allows users to simultaneously search HSDB, TOXLINE, CCRIS, Gene-Tox, DART/ETIC, IRIS, TRI and ChemIDplus from one input screen. Based on recommendations from the Institute of Medicine, users are presented with a basic search screen with just a single input box for searching, with customized screens for more sophisticated users. These advanced features include Boolean searching and the ability to limit search terms to specific fields. A TOXNET user online survey was carried out in the fall of 2001, and feedback from this survey is being used for current and future planning. New search screen designs were implemented in 2002, and research and development projects such as a chemical spellchecker, automatic indexing, and a toxicology gateway system were carried out.

Alternatives to Animal Testing—SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures that could be used to reduce,

refine, or replace animals in biomedical research and toxicological testing. Staff members search, edit, and categorize citations to create a true value added resource in this field. The 22 bibliographies issued during the past 10 years are available on the Internet through the SIS Web server, and the primary distribution mechanism for this project is now the Internet. In FY2002, a new online resource named ALTBIB was made public, allowing search access to all of the 7,595 citations organized from previous bibliographies. This uses the TOXNET search engine, and is available at (http://toxnet.nlm.nih.gov/altbib.html). A user may search by keyword, author, or one of the 16 subdivisions such as "Quantitative Structure Activity Studies."

TOXLINE (Toxicology Information Online) is a large NLM bibliographic database traditionally produced by merging "toxicology" subsets from secondary sources. By the end of FY2002, the database included over 3 million citations to toxicology literature going back to 1965. In FY2002, we completed the transition to a next generation TOXLINE, reducing the components needed to produce the database by creating a toxicology subset on NLM's PubMed so that users can access standard journal literature in toxicology and environmental health as part of an enlarging MEDLINE database. NLM added journals in the area of toxicology and environmental health to MEDLINE to cover some of the literature formerly provided by outside sources. For the non-standard journal literature in this area we created a Web-based system on TOXNET that allows efficient acquisition and updating of these components. Easy access to this TOXLINE Special database and to TOXLINE Core, the standard journal literature on PubMed, is available from the new TOXNET user interface.

DIRLINE (Directory of Information Resources On-line) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users which may not be available from one of the other NLM bibliographic or factual databases. DIRLINE continues to receive a high level of use through a new interface, which became public in October 1999. This new interface supports direct links to the Web sites of the organizations listed in the database, as well as direct e-mail connections. The quality and utility of the database continue to improve as duplicates have been

eliminated through changes in policy and streamlining of maintenance. Health Hotlines, the always popular publication of health-related toll-free telephone numbers, has a Web version that also lists Spanish speaking customer service representatives and Spanish language publications.

The **Toxics Release Inventory (TRI)** series of files now includes five online files, TRI95 through TRI2000. These files remain an important resource for environmental release data and are a useful complement to our other databases. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA databases contain environmental release data for air, water, and soil for over 600 EPA-specified chemicals. These files are used in the new SIS R&D project using a geographical information system, TOXMAP.

The **Chemical Carcinogenesis Research Information System** (CCRIS) continues to be built, maintained, and made publicly accessible at NLM. This data-bank is supported by the National Cancer Institute and has grown to over 8,000 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion and tumor inhibition.

The **Integrated Risk Information System** (IRIS), EPA's official health risk assessment file, continues to experience high usage and be very popular with the user community. EPA has had a version of IRIS on the agency's Web page since 1996, and we will continue to consider how best to integrate our Web service with what EPA provides. IRIS now contains 538 chemicals.

The **GENE-TOX** file is built directly on TOXNET by EPA scientific staff. This file contains peer-reviewed genetic toxicology (mutagenicity) studies for about 3,200 chemicals. GENE-TOX receives a high level of interest among users in other countries.

The **Registry of Toxic Effects of Chemical Substances** (RTECS) is a data-bank based upon a National Institute for Occupational Safety and Health (NIOSH) file by the same name which NLM restructured and made available for on-line searching. With our move to free Internet access to all databases, NIOSH requested that we no longer include RTECS on our system. We continue to use RTECS in the creation of the Hazardous Substance Data Bank.

The **Developmental and Reproductive Toxicology** (DART) database now contains over 240,000 citations from literature published since 1989 on agents that may cause birth defects. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database. In FY2002, we completed the transition to a next generation DART and created two subsets: DART Core on PubMed, containing over 170,000 citations to the journal literature and DART Special containing nearly 70,000 citations to specialized resources (including meeting abstracts, books, technical reports). Easy access to DART Special and to DART Core, is available from the new TOXNET interface. DART is funded by NLM, the EPA, the National Institute of Environmental Health Sciences (NIEHS), and the FDA's National Center for Toxicological Research, and is managed by NLM.

The **Environmental Mutagen Information Center** (EMIC) database contains over 24,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950 to 1991. The EPA, NIEHS, and NLM, collaborating partners in this effort, stopped compiling this special collection as of December 1999, but SIS will keep the collections as part of the TOXLINE Special database on TOXNET.

On March 21, 2002, SIS sponsored a Children's Environmental Health Information Resources Satellite Broadcast via the CDC Public Health Training Network. The program demonstrated selected online resources in the context of important children's environmental health issues. Topics included exposure of children to pesticides, environmental triggers of childhood asthma, methylmercury and fish contamination, the use of Geographic Information Systems for environmental health data, Healthy People 2010 resources, and lead poisoning prevention funding resources. The program was designed for physicians, nurses, physician assistants, nurse practitioners, epidemiologists, public health educators, librarians, counselors, administrators, or anyone else providing environmental health-related services. The Web cast of the broadcast is available at: (http://www.phppo.cdc.gov/PHTN//Webcast/child-env/archivewc.asp). In addition, a children's environmental health resource sampler was developed (http://nnlm.gov/partners/children/sampler.html).

AIDS Information Services

NLM has continued its successful AIDS Community Information Outreach Program with 15 new awards in FY2002, bringing the total number of awards made to 157. In addition to these awards, NLM has been working with other organizations to raise awareness of HIV/AIDS information resources among small community organizations at a grassroots level.

NLM remains as the project manager for the multi-agency AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS). These are in the process of being merged into a new service currently titled "AIDS info". This new service will continue to provide access to AIDS-related clinical trials information (through ClinicalTrials.gov) and federally approved treatment guidelines. The contract for this service also provides support services for ClinicalTrials.gov.

Outreach / User Support

Special Population Web Sites: The Arctic Health Web site (http://arctichealth.nlm.nih.gov), initially developed by SIS staff, has been turned over to the University of Alaska, Anchorage for continued development. This will remain a collaborative project between SIS, the Consortium Library and the Institute for Circumpolar Health Studies at the University. A users council has been established, which includes representatives from many of the stakeholder groups. Work is continuing on developing an Asian American Web site and an American Indian Web site. These Web sites include relevant policy, legislative, and organizational information as well as organized links to health and environmental issues of that particular population.

SIS collaborated in a training project with the DHHS Office of Minority Health. As part of their AIDS initiative, the Office conducted a needs assessment of community organizations in six major cities. Among the top needs identified by these community-based organizations was training in the use of the Internet to find health information resources. NLM provided training in identifying and using HIV/AIDS information resources for representatives of community-based organizations in six cities across the country.

SIS continues its support of the Toxicology Information Outreach Project (TIOP). The objective of this initiative is to strengthen the capacity of Historically Black Colleges and Universities (HBCUs) to train medical and other health professionals in the use of NLM's toxicological, environmental, occupational health and hazardous waste information resources. This year the panel held its annual meeting at the University of Puerto Rico Medical Science Campus. The University of Puerto Rico is one of the two new members of the panel. The assessment of the program was presented to the panel at their annual meeting. The panel has recommended expanding the scope of their activities beyond toxicology and environmental health, to encompass the health disparities that have been identified as disproportionately affecting minority communities.

SIS developed a new program of outreach to HBCUs in conjunction with the United Negro College Fund Special Programs Corporation (UNCFSP). This program gives NLM the opportunity of working with additional HBCUs that may not have graduate health programs. Technical assistance in the form of training in using health information resources will be provided, as will other forms of support and assistance. UNCFSP serves as the intermediary in recruiting HBCUs to participate and selecting projects for funding. Several TIOP representatives are serving on the advisory board for UNCFSP as well as serving as reviewers for funding awards.

SIS initiated a new Information Internship program this year. This internship was jointly funded by NLM and the National Center for Minority Health and Health Disparities. Two representatives from the Mandan, Hidatsa, and Arikara People (Three Affiliated Tribes) started a one-year program of working with NLM and learning about health information resources and access. The internship culminates in the development of a local project on their reservation intended to improve access to health information for the tribe. This first project will be the development of a mobile computer training facility that will be moved to different locations on the reservation and for classes and opportunities to access health information.

SIS undertook a major training program with the National Medical Association. In addition to providing training at the NMA Annual Meeting and participating in the sessions of the Community

Medicine Section, NLM provided day-long training sessions at six regional NMA meetings. These courses covered all of NLM's online resources including TOXNET, PubMed, ClinicalTrials.gov, and MEDLINEplus.

SIS participated in a health pilot project of the Department of Housing and Urban Affairs Neighborhood Network Centers. Neighborhood Network Centers with computers and Internet connectivity are available in approximately 1,000 HUD assisted multi-unit dwellings. During the pilot project health programs were held at 12 sites. NLM supplied a staff person to demonstrate retrieval of relevant information as part of each health program. In addition, a connection was made between the Center manager and local public and health sciences librarians.

A recent addition to NLM's outreach programs is one to improve access to health-related disaster information in three disaster-prone Central American countries: Nicaragua, Honduras, and El Salvador. NLM is funding the Regional Disaster Information Center for Latin America and the Caribbean (CRID) to strengthen the capacity of these countries to collect, index, manage, store, and disseminate public health and medical information related to disasters. The main objective of this project is to contribute to disaster reduction by capacity building activities in the area of disaster-related information management. Selected libraries and information centers have been provided with the knowledge, training and technology resources in order to act as reliable information providers to health professionals and others in their countries. Through this initiative, the participating libraries and information centers have been strengthened in several areas:

- Technological infrastructure (Internet connectivity and computer equipment)
- Information management (health science librarian training)
- Information product development (Digital Library, Web sites)

This project is also assisting SIS in developing models for collecting and exchanging health information in geographically isolated and disaster-prone environments and for handling non-traditional or unpublished literature, in this case on the health aspects of disasters.

SIS exhibited at over 30 conferences in this fiscal year. Several of these provided opportunities for presentations or workshops about NLM's information resources. In addition, SIS provided support for some conferences, including the HRSA Conference on American Indians/Alaska Natives—HIV/AIDS and Substance Abuse, the DHHS Office of Minority Health's National Leadership Summit on Eliminating Racial and Ethnic Disparities in Health, and the Symposium on Career Opportunities in Biomedical Sciences sponsored by the Association of Minority Health Professions Schools. NLM also sponsored the e-health track at the Technology Partnerships Conference held at the Georgia Centers for Advanced Telecommunications and Technology in Atlanta.

Research and Development Initiatives

To meet the mission of providing information on toxicology, environmental health, and targeted biomedical topics to the world, SIS has been developing new ways of presenting the world of hazardous chemicals in our environment to a wider audience. Projects include the following projects:

Household Products Ingredients Database: a Web resource for consumers that links brand name household products with their ingredient chemicals and potential adverse health effects. This pilot database will be ready for beta testing in early FY2003.

ToxTown: (http://toxtown.nlm.nih.gov/): a pilot project that explores how best to provide environmental health information to a general audience. ToxTown offers a graphic view of a typical town and points out environmental hazards that may be in that town. Users can click on a town location, like the school, and see a cutaway view of that building. Toxic chemicals that might be found in the school are listed, along with links to selected Internet resources about school environments. ToxTown will become available to the public from the SIS Web site in October 2002. Plans for an urban view and a rural setting are under way.

TOXMAP: a prototype system that uses maps of the United States to help users visually view data about chemicals released into the environment. It integrates data from the EPA's Toxic Release Inventory (TRI) with information about health effects, research citations, etc found in TOXNET

databases. Users can create nationwide or local area maps that show where chemicals are released into the air, water, and ground. TOXMAP also integrates data from other sources, such as demographic data from Census Bureau. TOXMAP provides region-specific links to chemical and bibliographic information. A beta test version is scheduled for release in the first quarter of FY2003

HSDB-in-the-Palm: The objective of this new initiative is to provide critical chemical information quickly and conveniently on a Personal Digital Assistant (PDA) for use by emergency responders (first 24 hours in hot-zone). The application is being developed in partnership with the Agency for Toxic Substances and Disease Registry's.

ToxPortal: a virtual meta-search tool for simultaneous searching of target information systems, displaying search results from targeted systems, and harvesting related concepts. The tool can be configured to define a set of target information/search tools, including SIS databases and searchable resources on the Web. Testing of the prototype is under way and a beta version will be ready for public release in FY2003.

Chemical Spellchecker: Spelling errors in user queries, especially in chemical name searches, and the lack of semantic query assistance are known shortcomings of our present retrieval system. A prototype spellchecker was developed in FY2002, and it incorporates a general dictionary, medical dictionary, and chemical dictionary, as well as a UMLS enhanced lexicon and English language grammar parser. The new spellchecker will be integrated into the TOXNET search engine in FY2003.

In these and other new initiatives, SIS continues to search for new ways to be responsive to user needs in acquiring and using toxicology and environmental health, HIV/AIDS, and other specialized information resourcs.